

DOI: 10.5846/stxb202104271109

杨海乐 张辉 吴金明 李君轶 王成友 杜浩 危起伟.模拟计算平行样对流域生物信息流估算的影响.生态学报 2022 42(13) : 5211–5225.

Yang H L , Zhang H , Wu J M , Li J Y , Wang C Y , Du H , Wei Q W. Simulating the impacts of parallel samples on the estimations of upstream-to-downstream watershed biological information flow. Acta Ecologica Sinica 2022 42(13) : 5211–5225.

模拟计算平行样对流域生物信息流估算的影响

杨海乐 张 辉 吴金明 李君轶 王成友 杜 浩 危起伟*

中国水产科学研究院长江水产研究所 农业农村部淡水生物多样性保护重点实验室, 武汉 430223

摘要: 流域生物信息流是流域生态学研究中的重要内容,是流域生态系统中的物质输移和能量输移过程的信息标记,是用 eDNA 技术调查评估河流水体中物种组成空间特征的基础。估算流域生物信息流是流域生态系统过程研究和 eDNA 技术调查评估河流水体中物种组成空间特征的关键。在有限的调查采样中,平行样的数量如何影响流域生物信息流的估算,尚待解答。基于随机抽样调查的基本原理,提出假设——采样数量不影响流域生物信息流估算结果的准确度,但会影响其精密度,然后通过问题简化转化和模拟计算,对该假设进行了检验。模拟计算结果显示,随着样点生物信息检出度(平行样数量)的增大,流域生物信息流估算结果会从偏小逐渐靠近流域生物信息流实际值,同时其 99.9% 置信区间也逐渐集中于流域生物信息流实际值。即样点生物信息检出度(平行样数量)对流域生物信息流估算的准确度和精密度均有影响。在实际调查研究过程中,建议先在所研究区域对平行样数量和样点生物信息检出度的关系进行预评估,然后基于流域生物信息流估算可信度目标在正式实施方案中经济有效地设置平行样,基于多平行样调查结果估算流域生物信息流,再根据各样点生物信息检出状况对流域生物信息流估算结果进行后验评估。

关键词: 流域生物信息流; 环境 DNA; 平行样; 检出度; 流域生态学

Simulating the impacts of parallel samples on the estimations of upstream-to-downstream watershed biological information flow

YANG Haile , ZHANG Hui , WU Jinming , LI Junyi , WANG Chengyou , DU Hao , WEI Qiwei*

Key Laboratory of Freshwater Biodiversity Conservation , Ministry of Agriculture and Rural Affairs of China , Yangtze River Fisheries Research Institute , Chinese Academy of Fishery Sciences , Wuhan 430223 , China

Abstract: Watershed biological information flow (WBIF) is defined as the path , processes and control of biological information transport , exchange , interaction and feedback among different spaces and systems along with watershed ecosystem processes , and could be partly described as the land-to-river and upstream-to-downstream bioinformation transportation (including organisms , nucleic acids , peptides and other biomarkers) , which is driven by the hydrologic processes of watershed systems. The WBIF labels the transport of organic matter and energy. The WBIF integrates the ecological processes of environmental DNA (eDNA) , including the origin , state , transport , and fate of eDNA , and makes it possible that the species composition in river system is monitored and assessed using eDNA. The WBIF estimation is the key for watershed ecosystem processes studying and riverine biodiversity monitoring. However , in practice , the parallel samples in each sampling site always are limited. And how parallel samples would impact WBIF estimation is unknown. Based on the principles of stochastic sampling survey , we hypothesized that parallel samples would not impact the accuracy of the WBIF estimation , but affect the precision of the WBIF estimation. Then , we transformed this hypothesis into a set of

基金项目: 农业农村部财政专项“长江渔业资源与环境调查”(CJDC-2017-14); 中国水产科学研究院中央级公益性科研院所基本科研业务费专项(2020JBF01 2020TD08)

收稿日期: 2021-04-27; 网络出版日期: 2022-03-18

* 通讯作者 Corresponding author. E-mail: weiqw@ yfi.ac.cn

<http://www.ecologica.cn>

chinaXiv:202101.00041v2

formulas and tested it with a series of analog computation. Results showed that the number of parallel samples (efficiency of detection) affected both the accuracy and precision of the WBIF estimation. The optimal WBIF estimation was less than the actual WBIF in any condition. Along with the increase of parallel samples (efficiency of detection) , the optimal WBIF estimation gradually neared to the actual WBIF , the range of WBIF estimation gradually focused on the actual WBIF. In other words , more parallel samples (higher efficiency of detection) led higher accuracy and precision of the WBIF estimation. In addition , the actual WBIF affected both the accuracy and precision of the WBIF estimation too. Larger actual WBIF led higher accuracy and precision of the WBIF estimation. The relative relationship between the number of biological information types in upstream and downstream samples affected both the accuracy and precision of the WBIF estimation too. The accuracy and precision of WBIF estimation would be higher when the number of biological information types in upstream samples was more than those in downstream samples. So , we suggest that in the work of watershed ecosystem processes studying and riverine biodiversity monitoring , the relationship between parallel sample number and detection efficiency should be assessed , the suitable parallel sample number should be estimated based on the reliability target of WBIF estimation , the sampling program should be designed with suitable parallel samples , the WBIF should be estimated based on all parallel samples of each sampling site , at last the estimated results of WBIF should be re-evaluated according to the posterior probability of WBIF in different conditions. The current work provided the framework and methodology reference for the post-evaluation.

Key Words: watershed biological information flow; environmental DNA; parallel sample; detection efficiency; watershed ecology

流域生态学(watershed ecology) 是研究流域范围内陆地和水体生态系统相互关系的学科(全国科学技术名词审定委员会) ,其主体在于整合研究以流域生态系统过程为核心所关联起的流域内各相关子系统(subsystem) ^[1]。流域生态系统过程研究的核心工作是研究流域生态系统中依托于水循环过程的各个子系统之间及之内的物质流、能量流、信息流^[2]。流域生物信息流(watershed biological information flow) 是生物信息依托于流域生态系统过程在不同空间和系统之间进行传递、交流、作用、反馈的路径、过程与控制 , 主要关注生物体及生物物质所承载的生物信息在时空上的迁移扩散 , 以及与生物体及生物物质间相互作用相伴随的生物信息作用和反馈^[3]。任何生物能都依托于生物物质 , 而任何类型的生物物质都携带有其特定的信息标记 , 因而作为流域生态系统过程研究的三大主题(物质流、能量流、信息流) 之一 , 流域生物信息流通过物质流、能量流、信息流的三位一体而具有对流域生态系统中的物质输移和能量输移过程的指征和标记功能 , 流域生物信息流的研究可以为流域生态系统研究中生物物质输移和生物能输移过程的跟踪和标记提供理论的可能和技术的支撑^[4-5]。

河流上游到下游的流域生物信息流是 eDNA(environmental DNA) 技术调查评估河流水体中物种组成空间特征的事实基础^[5-8] , 估算河流上游到下游的流域生物信息流是用 eDNA 技术调查评估河流水体中物种组成空间特征的方法基础^[9-11]。基于集合生态系统概念框架 根据研究尺度和分辨率的需要 , 可以将流域生态系统构建为一个包含了多个子系统的集合生态系统(meta-ecosystem) ^[12]。评估河流水体中物种组成的空间特征 , 即评估河流水体(基于空间划分) 各子系统内的物种组成。假设各子系统在特定时间的物种组成相对稳定 , 各子系统内的个体释放到环境中的 eDNA 则随着河川径流向下游输移。因为流水水体中的 eDNA 受释放、稀释、吸附、再悬浮、输移、降解等过程的影响^[13-15] , 上游水体子系统中个体释放到环境中的 eDNA 随着河川径流向下游输移的距离有限^[16-18] , 因而可以通过估算流域生物信息流来评估河流水体各子系统内的物种组成 , 进而评估河流水体中物种组成的空间特征。

分析流域生物信息流 , 第一步是估算上游子系统到下游子系统的流域生物信息流输移有效度(即用下游

样点水体 eDNA 监测上游样点水体中生物组成信息的监测有效度,上游样点水体中的水生生物信息能在下游样点水体的 eDNA 中被检出的比例和概率^[3]。监测有效度估算的核心是(1)有限采样对采样区域的生物组成信息的检出度和(2)由上游到下游的流域生物信息流输移有效度。在实践应用中,调查采样往往是有限的,因而存在两个问题:(1)平行样的设置情况如何影响生物组成信息检出度的估算,(2)平行样的设置情况如何影响流域生物信息流输移有效度的估算。基于抽样调查的基本原理和前期 eDNA 调查研究可知,采样数量越多对采样区域的生物组成信息的检出度越大,因检出度受多种因素影响,具体的平行样数量和检出度之间的关系需要一系列具体研究来进行量化^[19]。如果将由上游到下游的流域生物信息流过程简化为上游区域生物信息通过随机取样然后转移到下游区域,那么基于有限采样的流域生物信息流估算就是对随机抽样调查结果的随机抽样调查,基于随机抽样调查的基本原理可以猜测,采样数量不影响抽样调查结果的准确度,但会影响其精密度。

根据流域生物信息流研究框架^[3],本文就平行样的设置对流域生物信息流估算的影响展开了模拟计算,以检验“在流域生物信息流估算中各样点平行样数量不影响抽样调查结果的准确度,但会影响其精密度”的假设,并探讨在流域生物信息流估算中各样点平行样数量(检出度)的增加或减少对流域生物信息流估算的具体影响,为以下游水体 eDNA 监测上游水体中物种组成信息的监测有效度评估方案提供指导和支持,推动 eDNA 技术在水生生物多样性调查监测中的应用和发展。

1 研究方法

1.1 基本假设

随机抽样过程中,抽样数量越多,抽样结果的集合对整个系统的反映程度越全面。在本研究中,抽样即 eDNA 采样,抽样数量即 eDNA 样本数,对整个系统的反映程度即 eDNA 对该样点水体生物信息种类的检出度。因此在模拟计算中,我们将平行样数量的增加转化为对该样点水体生物信息种类检出度的增加。

上下游样点间共有生物信息种类组成占上游样点总生物信息种类组成的比例,即上游到下游的流域生物信息流。因为上下游相邻样点间的流域生物信息流是整体流域生物信息流计算的基础单元,所以本模拟计算中流域生物信息流的估算用上下游相邻样点间的共有生物信息组成占上游样点总生物信息组成的比例估算简单指代。

1.2 条件设定

平行样数量的增加对流域生物信息流估算的影响,用生物信息检出度的增加对上下游样点间共有生物信息种类组成占上游样点总生物信息种类组成的比例估算的影响来等价模拟展示。在本模拟计算中,设定上下游样点分别有生物信息 1000 种、900 种(可颠倒替换),样点生物信息检出度和实际流域生物信息流作为两个自变量,估算流域生物信息流为因变量,进行单变量模拟,具体分组及参数如表 1,然后计算特定上下游样点生物信息种类数组合情况下,样点生物信息检出度变化对估算流域生物信息流的影响(即实际流域生物信息流与估算流域生物信息流的差异)以及不同实际流域生物信息流条件下估算流域生物信息流的结果偏差程度。

1.3 计算方法

模拟计算的基本思路是,在特定上下游样点生物信息种类数组合情况下、在特定样点生物信息检出度情况下、在特定实际流域生物信息流情况下,通过随机取样计算得出某流域生物信息流值的概率,然后通过统计展示估算流域生物信息流相对于实际流域生物信息流的偏差,进而评估样点生物信息检出度变化对估算流域生物信息流的影响以及不同实际流域生物信息流条件下估算流域生物信息流的结果偏差程度。具体模拟计算公式如下列公式组

表 1 不同样点生物信息检出度条件下的流域生物信息流估算条件分组及参数设置

Table 1 Groups and parameters setting for watershed biological information flow (WBIF) estimation in different biological information detection efficiency (BIDE) conditions at sampling sites

分组 Groups	上游样点生物 信息种类数 Biological information types at up-site(M)	下游样点生物 信息种类数 Biological information types at down-site(N)	样点生物信息 检出度 BIDE at sampling sites(r)	实际流域生物 信息流 Actual WBIF(k)	估算流域生物 信息流 Estimated WBIF(e)
A	1000	900	x	0.5	y
B	1000	900	x	0.8	y
C	1000	900	0.1	x	y
D	1000	900	0.5	x	y
E	900	1000	x	0.5	y
F	900	1000	0.5	x	y

A、B、C、D、E、F 为分组的顺序编号; x 自变量; y 因变量

$$f(a_M) = \left(\frac{C_{M \times (1-r)}^{(M \times k - a)} \times C_{M \times r}^a}{C_M^{M \times k}} \right) \quad (a \leq M \times k; a \leq M \times r; 0 \leq r \leq 1; 0 \leq k \leq 1)$$

$$f(a_N) = \left(\frac{C_{N \times (1-r)}^{(M \times k - a)} \times C_{N \times r}^a}{C_N^{M \times k}} \right) \quad (a \leq M \times k; a \leq N \times r; 0 \leq r \leq 1; 0 \leq M \times k \leq N)$$

$$f(y) = \left(f(a_M) \times \sum_a^{M \times r} f(a_N) \right) + \left(f(a_N) \times \sum_a^{M \times r} f(a_M) \right) - \left(f(a_M) \times \sum_a^{M \times r} f(a_N) \right) \times \left(f(a_N) \times \sum_a^{M \times r} f(a_M) \right)$$

$$(y = \frac{a}{M \times r}; a \leq M \times r)$$

式中 M ,上游样点生物信息种类数; N ,下游样点生物信息种类数; r 样点生物信息检出度; k 实际流域生物信息流; a 监测到的生物信息种类数; y 所估算的流域生物信息流值; $f(a_M)$,上游样点处监测到数量为 a 的生物信息种类数的概率; $f(a_N)$,下游样点处监测到数量为 a 的生物信息种类数的概率; $f(y)$,估算出流域生物信息流值为 y 的概率。

根据该公式组 ,按照上述参数设定分 6 组进行模拟计算 ,然后通过组内和组间模拟计算结果的对比分析 ,探讨(1) 样点生物信息检出度状况对流域生物信息流估算的影响(即实际流域生物信息流与估算流域生物信息流的差异) ;(2) 在特定样点生物信息检出度条件下 ,不同实际流域生物信息流状况对流域生物信息流估算的影响 ;(3) 在特定样点生物信息检出度和实际流域生物信息流条件下 ,上下游样点生物信息种类数相对关系对流域生物信息流估算的影响。

2 结果与分析

2.1 样点生物信息检出度状况对估算流域生物信息流的影响

模拟计算显示 ,上下游样点分别有生物信息 1000 种、900 种的条件下 ,在实际流域生物信息流水平在 0.5 的时候 (1) 随着样点生物信息检出度从 0.1 逐步增长到 0.9 ,估算流域生物信息流的最优估值从 0.48(偏离 4%) 逐步增长到 0.4978(偏离 0.44%) ,即随着样点生物信息检出度的增长估算流域生物信息流的最优估值逐渐趋近于实际流域生物信息流水平; (2) 随着样点生物信息检出度从 0.1 逐步增长到 0.9 ,估算流域生物信息流的 99.9% 置信区间逐渐从 0.34—0.59(最大偏离 32%) 收缩到 0.4822—0.51(最大偏离 3.56%) ,即随着样点生物信息检出度的增长估算流域生物信息流的估值区间逐渐集中于实际流域生物信息流水平(图 1) 。

在实际流域生物信息流水平在 0.8 的时候 (1) 随着样点生物信息检出度从 0.1 逐步增长到 0.9 ,估算流域生物信息流的最优估值从 0.79(偏离 1.25%) 逐步增长到 0.7978(偏离 0.275%) ,即随着样点生物信息检出度的增长估算流域生物信息流的最优估值逐渐趋近于实际流域生物信息流水平; (2) 随着样点生物信息检出度从 0.1 逐步增长到 0.9 ,估算流域生物信息流的 99.9% 置信区间逐渐从 0.67—0.86(最大偏离 16.25%) 收缩

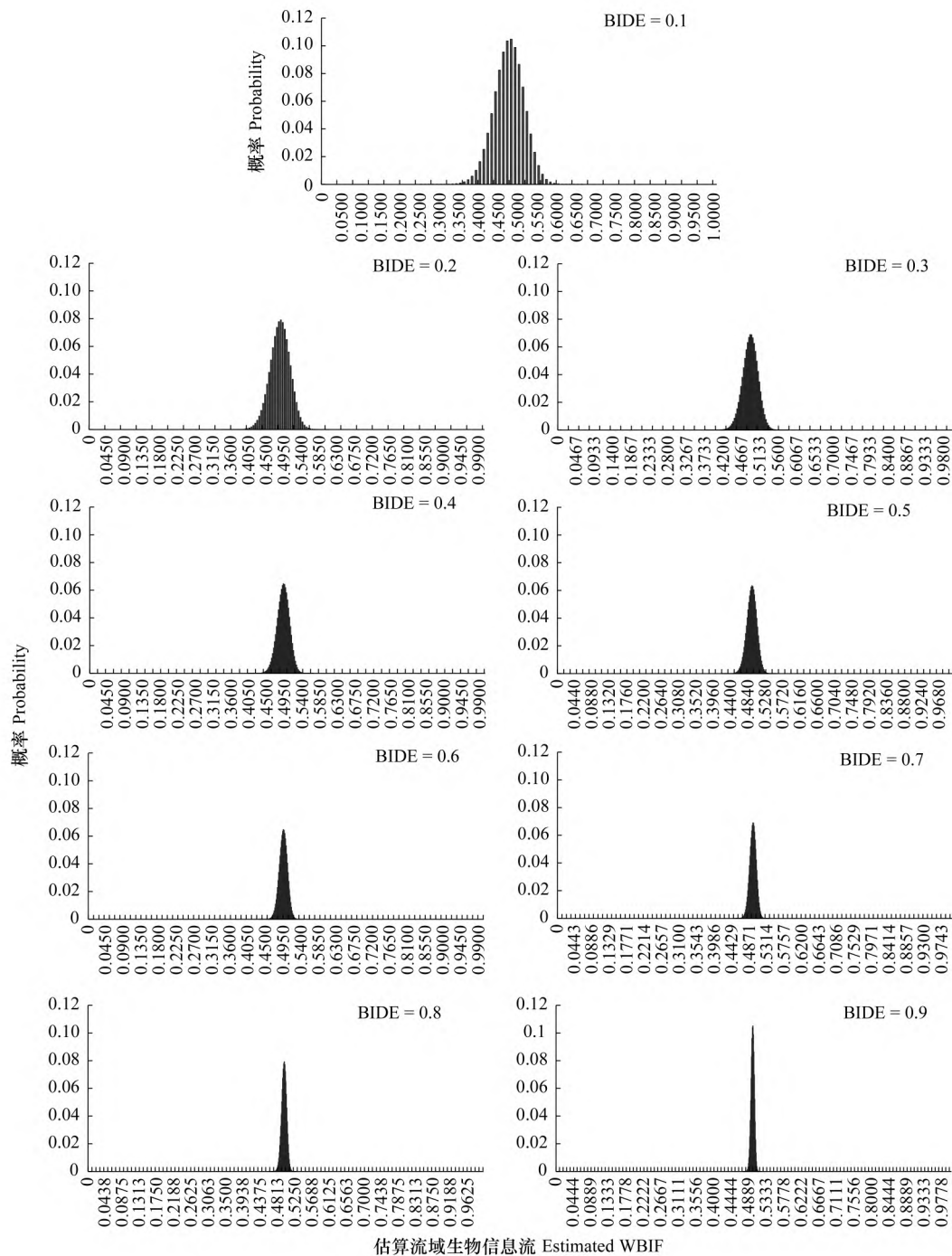


图 1 不同样点生物信息检出度下(A 组) 估算流域生物信息流值的概率分布

Fig.1 Probability distributions of estimated watershed biological information flow (WBIF) in different biological information detection efficiency (BIDE) conditions at sampling sites (group A)

A 组,上下游样点分别有生物信息 1000 种、900 种,实际信息流水平为 0.5,样点生物信息检出度为变量。WBIF: 流域生物信息流 watershed biological information flow; BIDE: 生物信息检出度 biological information detection efficiency

到 0.7867—0.8067(最大偏离 1.66%) ,即随着样点生物信息检出度的增长估算流域生物信息流的估值区间逐渐集中于实际流域生物信息流水平(图 2)。

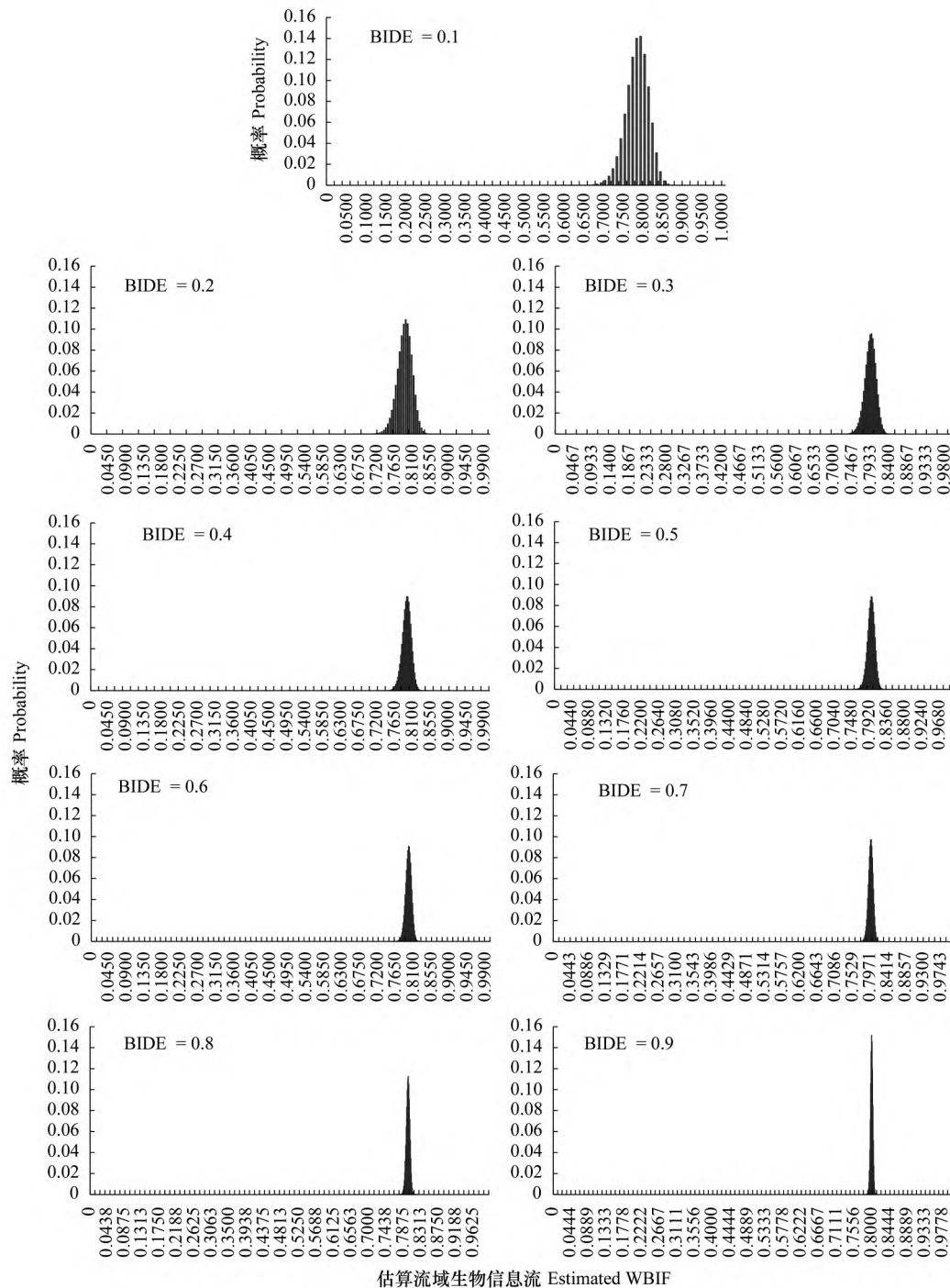


图2 不同样点生物信息检出度下(B组)估算流域生物信息流值的概率分布

Fig.2 Probability distributions of estimated watershed biological information flow (WBIF) in different biological information detection efficiency (BIDE) conditions at sampling sites (group B)

B组,上下游样点分别有生物信息1000种、900种,实际流域生物信息流水平为0.8,样点生物信息检出度为变量

2.2 实际流域生物信息流状况对估算流域生物信息流的影响

模拟计算显示,上下游样点分别有生物信息1000种、900种的条件下,在样点生物信息检出度为0.1时,(1)随着实际流域生物信息流从0.1增长到0.9,估算流域生物信息流的最优估值基本稳定在偏小0.02左右,但其偏离程度在逐步减小(从偏离20%到偏离趋近于0%),即随着实际流域生物信息流的增长估算流域生物

信息流的最优估值对实际流域生物信息流的偏离程度逐渐减小; (2) 随着实际流域生物信息流从 0.1 增长到 0.9, 估算流域生物信息流的 99.9% 置信区间幅宽逐渐先增大再减小(从 0.14(0.02—0.16) 逐步增大到 0.25(0.34—0.59) 再逐步缩小到 0.1(0.8—0.9)), 最大幅宽出现在实际流域生物信息流为 0.5 的时候, 但其最大偏离程度一直在逐步减小(从最大偏离 80% 到最大偏离 11%), 即随着实际流域生物信息流的增长估算流域生物信息流的估值区间幅宽先增大再减小, 但整体上逐渐相对集中于实际流域生物信息流水平(图 3)。

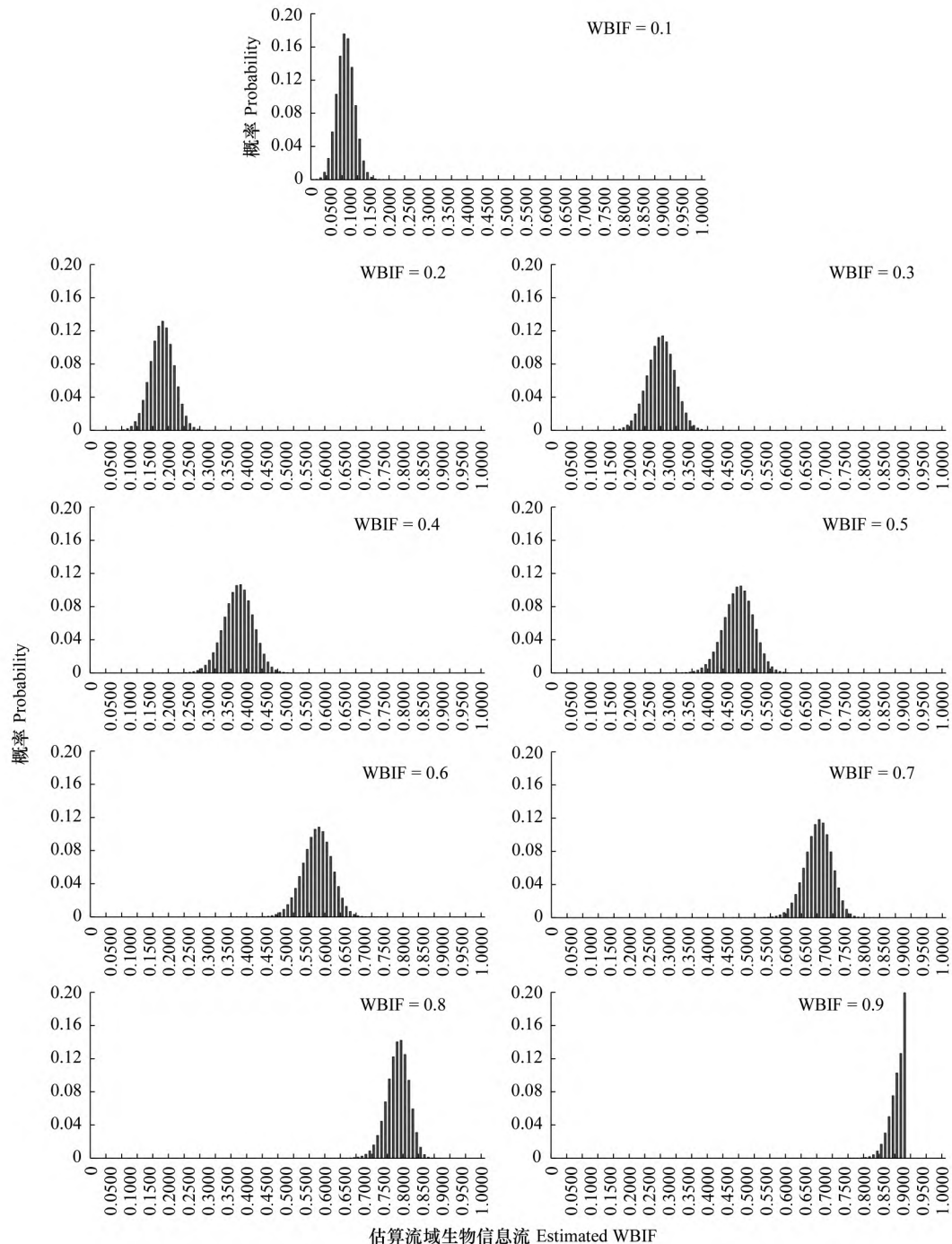


图 3 不同实际流域生物信息流下(C 组) 估算流域生物信息流值的概率分布

Fig.3 Probability distributions of estimated watershed biological information flow (WBIF) in different actual WBIF conditions (group C)

C 组, 上下游样点分别有生物信息 1000 种、900 种 样点生物信息检出度为 0.1 实际流域生物信息流为变量

在样点生物信息检出度为 0.5 时 (1) 随着实际流域生物信息流从 0.1 增长到 0.9, 估算流域生物信息流的最优估值从偏小 0.004 增大到偏小 0.008 再减小到 0, 但其偏离程度在逐步减小(从偏离 4% 到偏离趋近于 0%), 即随着实际流域生物信息流的增长估算流域生物信息流的最优估值对实际流域生物信息流的偏离程度逐渐减小; (2) 随着实际流域生物信息流从 0.1 增长到 0.9, 估算流域生物信息流的 99.9% 置信区间幅宽逐渐先增大再减小(从 0.05(0.068—0.118) 增大到 0.084(0.446—0.53) 再缩小到 0.03(0.87—0.9)), 最大幅宽出现在实际流域生物信息流为 0.5 的时候, 但其最大偏离程度一直在逐步减小(从最大偏离 32% 到最大偏离 3.3%), 即随着实际流域生物信息流的增长估算流域生物信息流的估值区间幅宽先增大再减小, 但整体上逐渐相对集中于实际流域生物信息流水平(图 4)。

2.3 上下游样点生物信息种类数相对关系对估算流域生物信息流的影响

模拟计算显示, 上下游样点生物信息种类数对调(分别为 900 种、1000 种)之后, 在实际流域生物信息流水平在 0.5 的时候 (1) 随着样点生物信息检出度从 0.1 逐步增长到 0.9, 估算流域生物信息流的最优估值从 0.4778(偏离 4.44%) 逐步增长到 0.4975(偏离 0.5%), 即随着样点生物信息检出度的增长估算流域生物信息流的最优估值逐渐趋近于实际流域生物信息流水平, 但相比于上下游样点生物信息种类数对调之前(上游 1000 种, 下游 900 种)的最优估值的偏离程度有所增大; (2) 随着样点生物信息检出度从 0.1 逐步增长到 0.9, 估算流域生物信息流的 99.9% 置信区间逐渐从 0.3222—0.6(最大偏离 35.56%) 收缩到 0.4802—0.5111(最大偏离 3.96%), 即随着样点生物信息检出度的增长估算流域生物信息流的估值区间逐渐集中于实际流域生物信息流水平, 但相比于上下游样点生物信息种类数对调之前(上游 1000 种, 下游 900 种)的估值区间的幅宽和最大偏离程度都有所增大(图 5)。

模拟计算显示, 上下游样点生物信息种类数对调(分别为 900 种、1000 种)之后, 在样点生物信息检出度为 0.5 的时候 (1) 随着实际流域生物信息流从 0.1 逐步增长到 0.9, 估算流域生物信息流的最优估值从偏小 0.0044(偏离 4.40%) 逐步增长到 0.0089(偏离 2.97%) 然后再缩小到 0.0067(偏离 0.74%), 即随着实际流域生物信息流的增长估算流域生物信息流的最优估值对实际流域生物信息流的偏离程度逐渐减小, 但相比于上下游样点生物信息种类数对调之前(上游 1000 种, 下游 900 种)的最优估值的偏离程度有所增大; (2) 随着实际流域生物信息流从 0.1 逐步增长到 0.9, 估算流域生物信息流的 99.9% 置信区间幅宽从 0.0543(最大偏离 33.3%) 逐渐先增大到 0.0934(最大偏离 11.56%) 再减小到 0.0666(最大偏离 4.93%), 即随着实际流域生物信息流的增长估算流域生物信息流的估值区间幅宽先增大再减小, 但整体上逐渐相对集中于实际流域生物信息流水平, 相比于上下游样点生物信息种类数对调之前(上游 1000 种, 下游 900 种)的估值区间的幅宽和最大偏离程度都有所增大(图 6)。

3 讨论

3.1 样点生物信息检出度同时影响了流域生物信息流估算的准确度和精密度

随着样点生物信息检出度的增长估算流域生物信息流的最优估值逐渐从偏低趋近于实际流域生物信息流水平。基于 A 组和 B 组参数的模拟计算结果显示, 上下游样点分别有生物信息 1000 种、900 种的条件下, 在实际流域生物信息流水平在 0.5 的时候, 随着样点生物信息检出度从 0.1 逐步增长到 0.9, 估算流域生物信息流的最优估值从 0.48(偏离 4%) 逐步增长到 0.4978(偏离 0.44%); 在实际流域生物信息流水平在 0.8 的时候, 随着样点生物信息检出度从 0.1 逐步增长到 0.9, 估算流域生物信息流的最优估值从 0.79(偏离 1.25%) 逐步增长到 0.7978(偏离 0.275%)。在样点生物信息检出度的各个水平条件下, 估算流域生物信息流的最优估值对实际流域生物信息流都偏小, 并且样点生物信息检出度越高估算流域生物信息流的最优估值对实际流域生物信息流的偏小程度越小, 估算准确度越高。

随着样点生物信息检出度的增长估算流域生物信息流的估值区间逐渐集中于实际流域生物信息流水平。基于 A 组和 B 组参数的模拟计算结果显示, 上下游样点分别有生物信息 1000 种、900 种的条件下, 在实际流

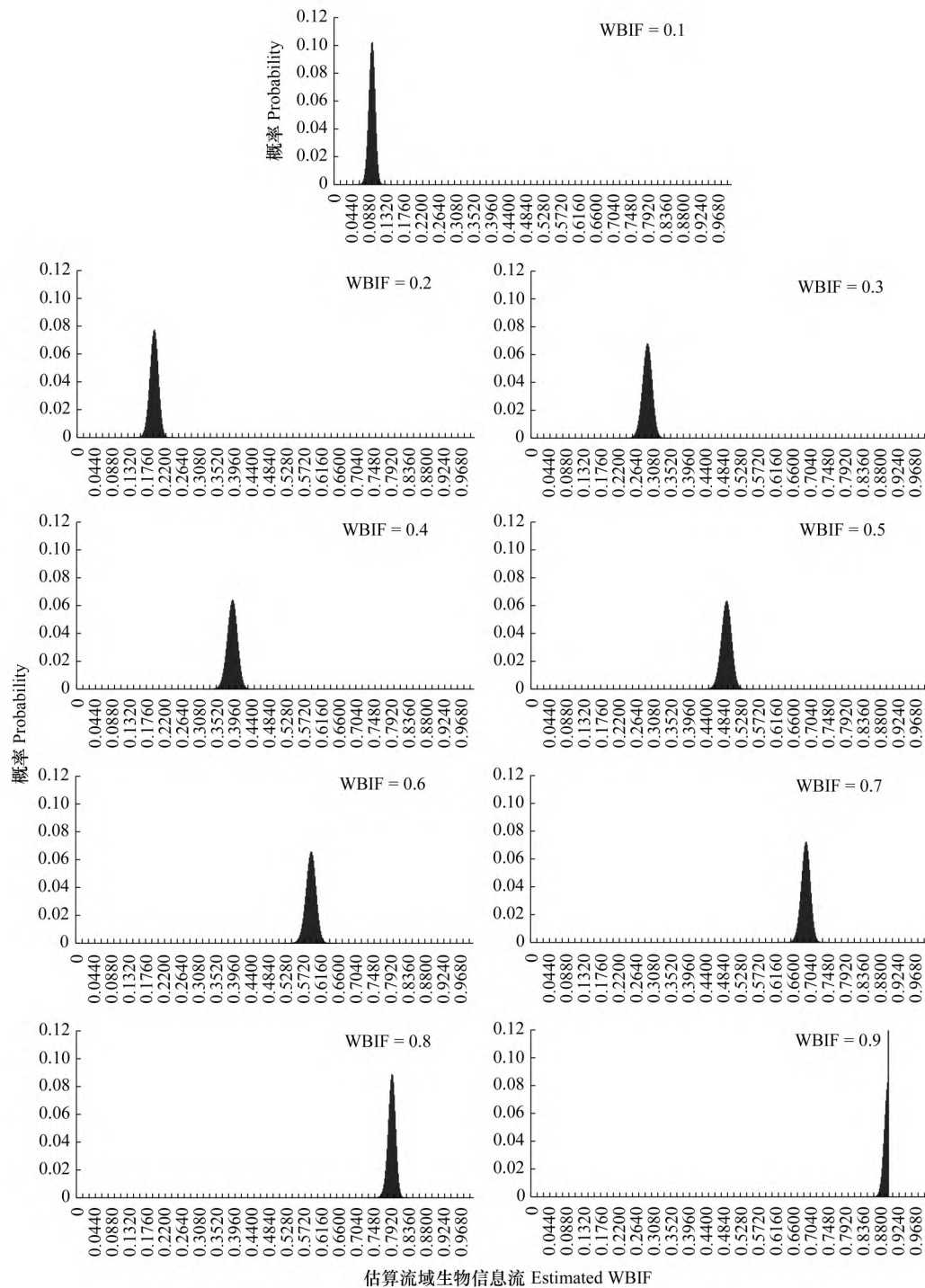


图4 不同实际流域生物信息流下(D组)估算流域生物信息流值的概率分布

Fig.4 Probability distributions of estimated watershed biological information flow (WBIF) in different actual WBIF conditions (group D)

D组,上下游样点分别有生物信息1000种、900种,样点生物信息检出度为0.5,实际流域生物信息流为变量

域生物信息流水平在0.5的时候,随着样点生物信息检出度从0.1逐步增长到0.9,估算流域生物信息流的99.9%置信区间逐渐从0.34—0.59(最大偏离32%)收缩到0.4822—0.51(最大偏离3.56%);在实际流域生物信息流水平在0.8的时候,随着样点生物信息检出度从0.1逐步增长到0.9,估算流域生物信息流的99.9%置信区间逐渐从0.67—0.86(最大偏离16.25%)收缩到0.7867—0.8067(最大偏离1.66%)。估算流域生物信息

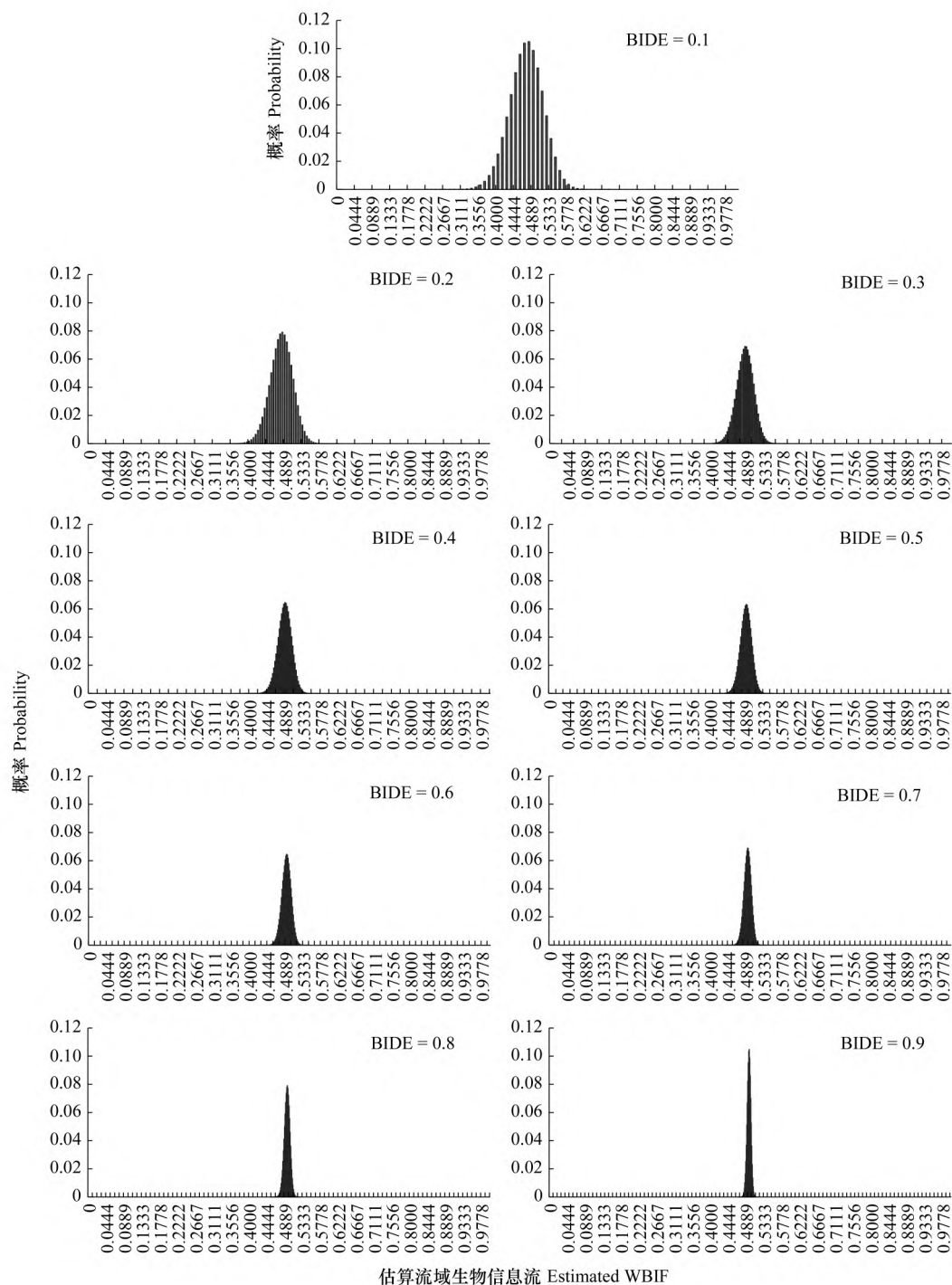


图5 不同样点生物信息检出度下(E组)估算流域生物信息流值的概率分布

Fig.5 Probability distributions of estimated watershed biological information flow (WBIF) in different biological information detection efficiency (BIDE) conditions at sampling sites (group E)

E组,上下游样点分别有生物信息900种、1000种,实际流域生物信息流水平为0.5,样点生物信息检出度为变量

流的99.9%置信区间内的概率分布近似正态分布,但向偏小的方向滑动(图1、图2),置信区间上限的最大偏离程度小于置信区间下限的最大偏离程度;样点生物信息检出度对估算流域生物信息流的99.9%置信区间内的概率分布对实际流域生物信息流的偏离程度有影响,检出度越高估算流域生物信息流的99.9%置信区间越小,对实际流域生物信息流的偏小程度也越小,估算精密度高。

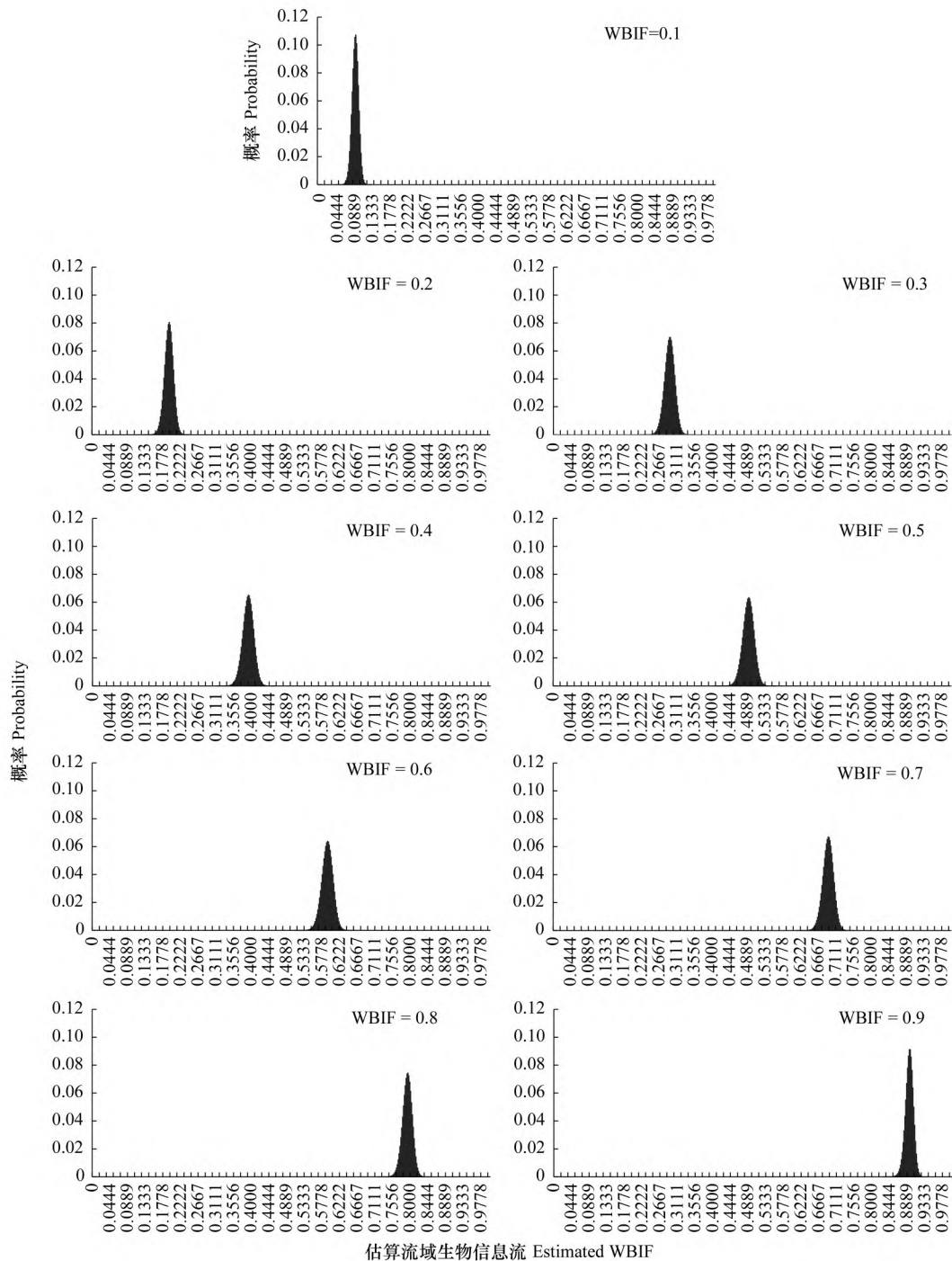


图 6 不同实际流域生物信息流下(F 组) 估算流域生物信息流值的概率分布

Fig.6 Probability distributions of estimated watershed biological information flow (WBIF) in different actual WBIF conditions (group F)

F 组 ,上下游样点分别有生物信息 900 种、1000 种 样点生物信息检出度为 0.5 实际流域生物信息流为变量

样点生物信息检出度越高,对样点生物信息的刻画越精确,对基于样点生物信息检出度的相关评估结果的估算也越精确。调查样点的生物信息检出是整个研究江段流域生物信息流估算的基础^[3],决定了用 eDNA 监测水生生物多样性的可监测距离的估算^[11]。前人的研究中,对流水水体中的水生生物的可监测距离的估算有各种各样的数值,从小几十米^[18-20]到几百米、数公里^[21-22]再到上百千米^[16]都有,导致这一结果差异的涉及到 eDNA 的产生、输移、吸附、沉积、再悬浮、降解等一系列动力学过程^[13-14],受各种各样的因素影

响^[17, 20, 23-24],也会有样点生物信息检出度差异对其估算结果的影响。为了推动基于 eDNA 技术的研究具有更好的可靠性和可重复性,标准化的研究方案和论文研究细节说明已被大家所注意到,但截至目前大家所关注的主要是采样点状况、采样技术方案、采样环境条件、采样时间设置、样品处理方式、污染控制状况、样品存储和运输条件、PCR 条件等^[25-28],较少关注平行样对研究结果影响的评估,以至于大家对用 eDNA 技术调查结果不做评估地直接展示^[29-30]。因此,本研究在此强调,对各 eDNA 监测研究所获得的可监测距离在采信之前需要给予相应基于样点生物信息检出度的可信度评估,同时也建议未来基于 eDNA 技术的研究对各自的研究结果进行可信度评估。

3.2 基于流域生物信息流估算可信度目标对样点生物信息检出度进行控制

为了得到可信的流域生物信息流估算结果,有必要基于流域生物信息流估算可信度目标进行样点生物信息检出度的控制。在实际调查应用中,调查到或没调查到某一目标对实际该目标存在与否的判断只是在一个概率意义上成立,当某一情况的可信度达到某一阈值其结果才可信或可用。与之相应,在流域生物信息流实际调查评估中,估算出的两个样点间的流域生物信息流值对实际流域生物信息流的表征也只是在一个概率意义上成立,当某一估算结果的可信度达到某一阈值其结果才可信或可用。假设我们设流域生物信息流估算的最优估值对实际值的偏离不超过 1%,且流域生物信息流估算值的 99.9% 置信区间对实际值的最大偏离不超过 5%,为流域生物信息流估算可信度的目标。那么,在上下游样点分别有生物信息 1000 种、900 种的,实际流域生物信息流水平在 0.5 左右的条件下,样点生物信息检出度应控制在 0.8 以上(图 1);在上下游样点分别有生物信息 1000 种、900 种的,实际流域生物信息流水平在 0.8 左右的条件下,样点生物信息检出度应控制在 0.5 以上(图 2)。

2020 年 9 月在武汉江段一个调查样点的 eDNA 调查结果显示,可用 eDNA (单引物 mlCOIintF/jgHCO2198R) 调查到的鱼类物种最优估计约 99 种,单样品 eDNA 的鱼类物种检出能力约为 26 种,检出效率约为 25.8%,在 50% 的检出度目标下,需要约 3 个平行样,在 80% 的检出度目标下,需要约 10 个平行样^[19]。如果在这一时段需要在长江干流武汉江段进行 eDNA 采样估算以鱼类为指标的流域生物信息流,其中假设采样距离设置为 10km,实际流域生物信息流预估为 0.8 左右,那么为了获得一个可信的流域生物信息流正式估算值,每个样点需采的平行样数量不应少于 3 个;假设如果采样距离设置为 30km,实际流域生物信息流预估为 0.5 左右,那么为了获得一个可信的流域生物信息流正式估算值,每个样点需采的平行样数量不应少于 10 个。

在两个采样点间的实际流域生物信息流值较低的情况下,应适当提高平行样数量。实际流域生物信息流大小不影响样点生物信息检出度对估算流域生物信息流的影响规律(图 1、图 2 对比),但影响了估算流域生物信息流的估算(图 3、图 4)。流域生物信息流本身的价值越低,采样和计算中的不确定性产生的影响就越大;流域生物信息流本身的价值越高,采样和计算中的不确定性产生的影响就越小。用模拟计算结果来讲,估算流域生物信息流的最优估值随着实际流域生物信息流降低而逐渐偏离(偏小)实际流域生物信息流水平(图 3),估算流域生物信息流的 99.9% 估值区间的最大偏离程度随实际流域生物信息流降低而逐渐增大,估值区间相对于实际流域生物信息流水平的集中度逐渐降低(图 4),所以在实际流域生物信息流的值较低时,应适当提高平行样数量,以通过提高样点生物信息检出度来提高估算流域生物信息流的精确度。2019 年在青藏高原一个小型河流的流域生物信息流量化研究结果显示,春季封冻期、夏季降雨天、秋季多云天的流域生物信息流估算结果分别为 75.86%、97.41%、96.07% 每千米^[11],如果以 10km² 为间隔设置样点,其样点间流域生物信息流估值分别为 6.31%、76.92%、66.97%,那么为了获得一个可信的流域生物信息流正式估算值,各样点的信息检出度分别需要达到 98%、50%、60% 左右。

在上游样点生物信息种类数小于下游样点生物信息种类数的情况下,应适当提高平行样数量。上游样点的生物信息种类与下游样点的生物信息种类的相对关系,并不影响样点生物信息检出度和实际流域生物信息流对估算流域生物信息流的影响规律,但影响了估算流域生物信息流的估算。由于在上游样点生物信息种类

数小于下游样点生物信息种类数的情况下所估算的流域生物信息流估值比在下游样点生物信息种类数小于上游样点生物信息种类数的情况下所估算的流域生物信息流估值相对流域生物信息流实际值有更大的最优估值偏离(偏小)程度(图 1、图 5 对比)和更大的 99.9%估值区间的幅宽和最大偏离程度(图 4、图 6 对比),所以在上游样点生物信息种类数小于下游样点生物信息种类数时,应适当提高平行样数量,以提高样点生物信息检出度来提高估算流域生物信息流的精确度。由于流域生物信息流会增加输入水体的生物信息种类^[3-4],因此通常情况下,上游样点的生物信息种类会少于下游样点的生物信息种类,在这种条件下,估算流域生物信息流最优估值对实际流域生物信息流的偏小程度更大,估算流域生物信息流估值区间的幅宽和最大偏离程度也更大。由于流域生物信息流对所输入水体的生物信息种类的增加在持续的向下游输移过程中是非累计的^[3],所以河流的干支流交汇处的生物信息种类会比交汇处下游较远处的生物信息种类要多,在这种条件下,估算流域生物信息流最优估值对实际流域生物信息流的偏小程度会减小,估算流域生物信息流估值区间的幅宽和最大偏离程度也会减小。另外,随着水体理化环境的转变,流域生物信息流中的有效流域生物信息流部分会被环境过滤效应所过滤掉,导致上游样点的生物信息种类会多于下游样点的生物信息种类^[3],在这种条件下,估算流域生物信息流最优估值对实际流域生物信息流的偏小程度会减小,估算流域生物信息流估值区间的幅宽和最大偏离程度也会减小。因而,相比于在河流交汇处及河流理化环境状况急剧转换处,在通常的河流条件下,流域生物信息流调查估算中需要更多的平行样。

3.3 对流域生物信息流估算结果进行后验概率评估

在上下游样点生物信息种类数相对关系确定、实际流域生物信息流确定、样点生物信息检出度确定的情况下,理论上所可得估算流域生物信息流的值依然是分布在某一特定估值区间,而非一个确切值。虽然说通过基于流域生物信息流估算可信度目标的样点生物信息检出度控制可以提高理论上所可得估算流域生物信息流的估值区间的集中度,进而使得流域生物信息流估算值对实际值有较好的近似程度,但在实际应用中还是有必要给出这个流域生物信息流估值所对应的实际流域生物信息流(区间)及相应后验概率,以保障其结果的科学性严谨性,给不同研究之间的横向比较提供科学可行合理可靠的基础。

在实际调查评估工作中,上下游样点生物信息种类数可以通过各调查样点一定数量的平行样的调查结果的物种积累曲线的分析而估算获得^[19],各调查样点的样点生物信息检出度可以通过调查所得的生物信息种类数除以各相应调查样点的生物信息种类数而计算获得^[19],估算流域生物信息流的值可以通过上下游共有生物信息种类数除以上游样点生物信息种类数而计算获得^[3],然后可以根据模拟计算公式组进行相应的模拟计算,再通过后验概率计算实际流域生物信息流状况及其概率分布。

流水水体中,自上游到下游的流域生物信息流过程十分复杂,涉及到 eDNA 的释放、稀释、吸附、再悬浮、输移、降解等过程^[13-15]。有一种方向是朝着弄清楚机理的方向努力,通过受控实验弄清环境因子对流域生物信息流过程中各环节的影响,再通过各环境因子的参数化、流域生物信息流过程各环境的模型化来推动对整个流域生物信息流的理解和量化^[17,31-33]。我们认为,还可以有另外一个方向,即暂时先把流域生物信息流的复杂过程打包成黑箱,接受不确定性控制不确定性量化不确定性,通过输入和输出的量化分析,推进流域生物信息流研究的应用和发展。这即是我们所提出的流域生物信息流分析框架的处理问题方式^[3,11],也是本研究的目的。

4 结论

本文针对平行样的增减是否会显著改变以及如何改变流域生物信息流的估算结果这一问题,通过对问题进行简化而转化为检出度是否会以及会如何影响估算流域生物信息流的问题,进而通过模拟计算对其进行了探讨。模拟计算结果显示,估算流域生物信息流的最优估值小于实际流域生物信息流,随着样点生物信息检出度(即平行样数量)的增长估算流域生物信息流的最优估值逐渐趋近于实际流域生物信息流值,估算流域生物信息流的估值区间逐渐集中于实际流域生物信息流值,即样点生物信息检出度越高(平行样越多)估

算流域生物信息流的准确度和精密度越高。另外,实际流域生物信息流大小、上下游样点生物信息种类数相对关系均对估算流域生物信息流的准确度和精密度有明确影响:实际流域生物信息流越大估算流域生物信息流的准确度和精密度越高,上游样点的生物信息种类多于下游样点的生物信息种类时的估算流域生物信息流的准确度和精密度相对更高。在实际应用中,为了提高流域生物信息流估算结果的可靠性,有必要基于流域生物信息流估算可信度目标进行样点生物信息检出度的控制,并且对流域生物信息流估算结果进行后验概率评估。

参考文献(References):

- [1] 杨海乐,陈家宽. 流域生态学的发展困境——来自河流景观的启示. 生态学报, 2016, 36(10): 3084–3095.
- [2] 蔡庆华,吴刚,刘建康. 流域生态学: 水生态系统多样性研究和保护的一个新途径. 科技导报, 1997, 15(5): 24–26.
- [3] 杨海乐,杜浩,祁洪芳,俞录贤,危起伟. 青藏高原沙柳河流域自然径流驱动的流域生物信息流量化特征——以环境微生物为指标. 生态学报, 2021, 41(9): 3475–3487.
- [4] Deiner K, Fronhofer E A, Mächler E, Walser J C, Altermatt F. Environmental DNA reveals that rivers are conveyor belts of biodiversity information. *Nature Communications*, 2016, 7: 12544.
- [5] Jerde C L, Olds B P, Shogren A J, Andruszkiewicz E A, Mahon A R, Bolster D, Tank J L. Influence of stream bottom substrate on retention and transport of vertebrate environmental DNA. *Environmental Science & Technology*, 2016, 50(16): 8770–8779.
- [6] Carraro L, Hartikainen H, Jokela J, Bertuzzo E, Rinaldo A. Estimating species distribution and abundance in river networks using environmental DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(46): 11724–11729.
- [7] Shogren A J, Tank J L, Egan S P, Bolster D, Riis T. Riverine distribution of mussel environmental DNA reflects a balance among density, transport, and removal processes. *Freshwater Biology*, 2019, 64(8): 1467–1479.
- [8] Jo T, Murakami H, Masuda R, Sakata M K, Yamamoto S, Minamoto T. Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources*, 2017, 17(6): e25–e33.
- [9] Wacker S, Fossøy F, Larsen B M, Brandsegg H, Sivertsgård R, Karlsson S. Downstream transport and seasonal variation in freshwater pearl mussel (*Margaritifera margaritifera*) eDNA concentration. *Environmental DNA*, 2019, 1(1): 64–73.
- [10] Sansom B J, Sassoubre L M. Environmental DNA (eDNA) shedding and decay rates to model freshwater mussel eDNA transport in a river. *Environmental Science & Technology*, 2017, 51(24): 14244–14253.
- [11] Yang H L, Du H, Qi H F, Yu L X, Hou X D, Zhang H, Li J Y, Wu J M, Wang C Y, Zhou Q, Wei Q W. Effectiveness assessment of using riverine water eDNA to simultaneously monitor the riverine and riparian biodiversity information. *Scientific Reports*, 2021, 11: 24241.
- [12] 杨海乐,陈家宽. 集合生态系统研究 15 年回顾与展望. 生态学报, 2018, 38(13): 4537–4555.
- [13] Barnes M A, Turner C R. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 2016, 17(1): 1–17.
- [14] Shogren A J, Tank J L, Andruszkiewicz E, Olds B, Mahon A R, Jerde C L, Bolster D. Controls on eDNA movement in streams: transport, retention, and resuspension. *Scientific Reports*, 2017, 7: 5065.
- [15] Cristescu M E, Hebert P D N. Uses and misuses of environmental DNA in biodiversity science and conservation. *Annual Review of Ecology, Evolution, and Systematics*, 2018, 49: 209–230.
- [16] Pont D, Rocle M, Valentini A, Civade R, Jean P, Maire A, Roset N, Schabuss M, Zornig H, Dejean T. Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports*, 2018, 8: 10361.
- [17] Seymour M. Rapid progression and future of environmental DNA research. *Communications Biology*, 2019, 2: 80.
- [18] Stoeckle B C, Kuehn R, Geist J. Environmental DNA as a monitoring tool for the endangered freshwater pearl mussel (*Margaritifera margaritifera* L.): a substitute for classical monitoring approaches? *Aquatic Conservation: Marine and Freshwater Ecosystems*, 2016, 26(6): 1120–1129.
- [19] 杨海乐,吴金明,张辉,杜浩,李君轶,王成友,沈丽,刘志刚,危起伟. 大型河流中鱼类组成的 eDNA 监测效率: 以长江武汉江段为例. *中国水产科学*, 2021, 28(6): 796–807.
- [20] Pilliod D S, Goldberg C S, Arkle R S, Waits L P. Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources*, 2014, 14(1): 109–116.
- [21] Tillotson M D, Kelly R P, Duda J J, Hoy M, Kralj J, Quinn T P. Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. *Biological Conservation*, 2018, 220: 1–11.
- [22] Deiner K, Altermatt F. Transport distance of invertebrate environmental DNA in a natural river. *PLoS One*, 2014, 9(2): e88786.

- [23] Barnes M A , Turner C R , Jerde C L , Renshaw M A , Chadderton W L , Lodge D M. Environmental conditions influence eDNA persistence in aquatic systems. *Environmental Science & Technology* , 2014 , 48(3) : 1819–1827.
- [24] Seymour M , Durance I , Cosby B J , Ransom-Jones E , Deiner K , Ormerod S J , Colbourne J K , Wilgar G , Carvalho G R , de Bruyn M , Edwards F , Emmett B A , Bik H M , Creer S. Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Communications Biology* , 2018 , 1: 4.
- [25] Goldberg C S , Turner C R , Deiner K , Klymus K E , Thomsen P F , Murphy M A , Spear S F , McKee A , Oyler-McCance S J , Cornman R S , Laramie M B , Mahon A R , Lance R F , Pilliod D S , Strickler K M , Waits L P , Fremier A K , Takahara T , Herder J E , Taberlet P. Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution* , 2016 , 7(11) : 1299–1307.
- [26] Dickie I A , Boyer S , Buckley H L , Duncan R P , Gardner P P , Hogg I D , Holdaway R J , Lear G , Makiola A , Morales S E , Powell J R , Weaver L. Towards robust and repeatable sampling methods in eDNA-based studies. *Molecular Ecology Resources* , 2018 , 18(5) : 940–952.
- [27] Nicholson A , McIsaac D , MacDonald C , Gec P , Mason B E , Rein W , Wrobel J , Boer M , Milián-García Y , Hanner R H. An analysis of metadata reporting in freshwater environmental DNA research calls for the development of best practice guidelines. *Environmental DNA* , 2020 , 2(3) : 343–349.
- [28] 邢迎春,高婉茹,白洁,赵亚辉. 环境 DNA 在湖泊生物多样性研究中的应用. *水生生物学报*, 2021: 1–20.
- [29] 徐念,熊美华,邵科,阙延福,李键庸. 长江中下游环境 DNA 宏条形码生物多样性检测技术初步研究. *环境科学研究*, 2020 , 33(5) : 1187–1196.
- [30] 凌建忠,姜亚洲,孙鹏,袁兴伟,张辉,唐保军. 环境 DNA 技术在象山港水域鱼类多样性调查中的应用与评估. *中国水产科学*, 2021 , 28(2) : 205–214.
- [31] 张娜,仇保丰,谢艳辉,李家桥,斯泽恩. 水环境生态系统中环境 DNA 脱落、衰变和运输过程的研究进展. *海洋湖沼通报*, 2021 , 43(3) : 129–135.
- [32] 李苗,单秀娟,王伟继,丁小松,戴芳群,吕丁,吴欢欢. 环境 DNA 在水体中存留时间的检测研究——以中国对虾为例. *渔业科学进展*, 2020 , 41(1) : 51–57.
- [33] Jo T , Arimoto M , Murakami H , Masuda R , Minamoto T. Particle size distribution of environmental DNA from the nuclei of marine fish. *Environmental Science & Technology* , 2019 , 53(16) : 9947–9956.